Electronic Version
Stylesheet Version v1.1.1

Description

FET GATE STRUCTURE WITH METAL GATE ELECTRODE AND SILICIDE CONTACT

BACKGROUND OF INVENTION

[0001] This application is related to Application FIS9-2003-0341, "Method of forming FET silicide gate structures incorporating inner spacers," filed the same day and assigned to the same assignee as the present application. The disclosure of the above-noted application is incorporated herein by reference.

[0002] This invention relates to the manufacture of advanced semiconductor devices, particularly advanced CMOS integrated devices in which metal gate electrodes are used.

[0003] With the continued scaling of CMOS devices to smaller dimensions, the gate dielectrics of these devices have been reduced to thicknesses well below 20 Å. This in turn has led to greatly increased gate leakage currents and diffusion of dopants from the polysilicon gate structures (often

referred to as the poly depletion effect).

[0004] Metal gates are now being used to mitigate the poly depletion effect and control the leakage current, and thus to ensure electrical performance in highly integrated CMOS devices. A metal gate is typically formed by a "replacement gate" process, in which a dummy polysilicon gate is first formed and then removed, and a metal gate formed in its place. The metal gate may span both n+ and p+ gate areas and comprise a single metal with a midgap work function. Alternatively, the replacement gate may comprise two metals with different work functions, in the space previously occupied by the n+ and p+ polysilicon gates respectively.

[0005] As an example of the present state of the art, U.S. Patent Application Publication No. 2003/0119292 of Lee et al., "Integration of dual workfunction metal gate CMOS devices," describes a replacement-gate process for forming a metal gate in which a doped polysilicon gate is formed and then removed, leaving an open trench; a bulk metal layer is deposited in the trench and then planarized to yield a metal gate. This approach may present manufacturing problems in two ways. First, a very narrow gate structure (corresponding to a very short channel length,

perhaps less than 70 nm) results in a high-aspect-ratio trench which may be difficult to fill without introducing a void in the metal. Second, a metal planarization process (typically chemical-mechanical polishing or CMP) is susceptible to dishing effects which cause nonuniformities in the metal thickness (and thus in the height of the metal gate).

[0006] It is also desirable to provide a silicided contact for the metal gate electrode. Accordingly, there is a need for a metal gate CMOS device with a silicide contact thereto, which is easily manufacturable and applicable to both a midgap replacement gate and a dual-metal replacement gate.

SUMMARY OF INVENTION

[0007] The present invention addresses the above-described need by providing a method for fabricating a semiconductor device having a gate structure on a substrate. According to a first aspect of the invention, the method begins with removal of material in a gate region of the device (that is, removal of a dummy gate structure and sacrificial gate dielectric) to expose a portion of the substrate. A gate dielectric is formed on the exposed portion of the substrate, and a metal layer is formed overlying the gate

dielectric and the dielectric material. This metal layer may conveniently be a blanket metal layer covering a device wafer. A silicon layer is then formed overlying the metal layer; this layer may also be a blanket wafer. A planarization or etchback process is then performed in which portions of the metal layer and the silicon layer are removed, so that the top surface of the dielectric material is exposed while other portions of the metal layer and the silicon layer remain in the gate region and have surfaces coplanar with the top surface of the dielectric material. A silicide contact is then formed which is in contact with the metal layer in the gate region.

[0008] The silicide contact may be formed by depositing a layer of a silicide-forming metal (Ni, Co, Ta, W or Mo) over the gate region; performing a siliciding process to form a metal silicide including silicon from the portion remaining in the gate region and the silicide-forming metal; and performing a planarization process to expose the top surface of the dielectric material.

[0009] The removal of the dummy gate material may be viewed as forming a trench, where the bottom is the exposed portion of the substrate. The step of forming the gate dielectric thus covers the bottom of the trench, while the

step of forming the metal layer thus forms metal on the sidewalls of the trench; the step of forming the silicon layer fills the trench. The silicon is afterwards converted to a silicide, so that the trench is filled with the silicide contact, the metal gate being a layer on the sidewalls and overlying the gate dielectric on the bottom of the trench.

[0010] According to a second aspect of the invention, two metal layers are formed in the gate region, so that a dual-metal gate structure is formed. In this method, material is removed from a first portion of the gate region, and afterwards from a second portion of the gate region, exposing portions of the substrate. The gate dielectric is formed on the exposed portions of the substrate. A first metal layer and a first silicon layer are formed overlying the gate dielectric and then planarized. A portion of the first metal layer is oxidized; this metal oxide layer separates the first and second metal layers. The second metal layer and second silicon layer are formed, and then planarized. A silicide contact is then formed in the gate region, in contact with both planarized portions of the first and second metal layers.

[0011] The silicide contact may be formed by forming a third silicon layer overlying both portions of the gate region; de-

positing a layer of a silicide-forming metal thereon; and performing a siliciding process. The siliciding process forms a metal silicide including silicon from remaining portions of the first and second silicon layers and from the third silicon layer, and metal from the layer of silicide-forming metal. A planarization process is then performed to expose the top surface of the dielectric material. The metal and metal oxide may be partially removed from the gate region so that the metal layers and oxide layer are recessed with respect to the top surface of the adjacent dielectric material. The third silicon layer then fills the recess, so that the silicide later likewise fills the recess.

[0012] According to another aspect of the invention, a semiconductor device having a gate structure on a substrate is provided. The gate structure of the device is fabricated as either a single-metal or dual-metal replacement gate, in accordance with the methods described above.

Brief Description of Drawings

[0013] Figure 1A is a schematic illustration in plan view of a typical CMOS dummy gate structure with n+ and p+ polysilicon regions.

[0014] Figure 1B is a longitudinal cross-section view of the gate regions of Figure 1AFigure 1C is a transverse cross-sec-

tion view of the gate regions of Figure 1A.

[0015] Figures 2A and 2B are longitudinal and transverse cross-section views, respectively, of a step in a midgap replacement gate process, in accordance with a first embodiment of the invention.

[0016] Figures 3A and 3B are longitudinal and transverse cross-section views, respectively, of a further step in a midgap replacement gate process, in accordance with the first embodiment of the invention.

[0017] Figures 4A and 4B are longitudinal and transverse cross-section views, respectively, of a further step in a midgap replacement gate process, in accordance with the first embodiment of the invention.

[0018] Figures 5A and 5B are longitudinal and transverse cross-section views, respectively, of an additional step in a midgap replacement gate process, in accordance with the first embodiment of the invention.

[0019] Figures 6-9 schematically illustrate steps in a dual-metal replacement gate process, in accordance with a second embodiment of the invention.

[0020] Figures 10A and 10B schematically illustrate further steps in a dual-metal replacement gate process, in accordance with the second embodiment of the invention.

[0021] Figures 11A and 11B are longitudinal and transverse cross-section views, respectively, of a further step in a dual-metal replacement gate process, in accordance with the second embodiment of the invention.

[0022] Figures 12A and 12B are longitudinal and transverse cross-section views, respectively, of an additional step in a dual-metal replacement gate process, in accordance with the second embodiment of the invention.

[0023] Figure13 illustrates a further step in a dual-metal replacement gate process, in accordance with the second embodiment of the invention.

[0024] Figures 14A and 14B are longitudinal and transverse cross-section views, respectively, of a further step in a dual-metal replacement gate process, in accordance with the second embodiment of the invention.

[0025] Figure15 illustrates an additional step in a dual-metal replacement gate process, in accordance with the second embodiment of the invention.

[0026] Figures 16A and 16B are longitudinal and transverse cross-section views, respectively, of another step in a dual-metal replacement gate process, in accordance with the second embodiment of the invention.

[0027] Figures 17A and 17B are longitudinal and transverse

cross-section views, respectively, of an additional step in a dual-metal replacement gate process, in accordance with the second embodiment of the invention.

[0028] Figures 18A-18C are transverse cross-section views of additional step in a dual-metal replacement gate process, as an alternative to the steps of Figures 17A and 17B.

DETAILED DESCRIPTION

[0029] The embodiments of the invention will be described as part of a replacement-gate process, in which a dummy polysilicon gate stack is built on a substrate and removed after source and drain regions are formed. Figures 1A-1C are schematic illustrations of the dummy gate structure, which is the starting point for the embodiments of the invention described herein. Figure 1A is a plan view of structure 10, which includes p+ polysilicon gate 11 and n+ polysilicon gate 12. As shown in Figure 1A, the n+ and p+ regions are in contact; this structure is typically found in SRAM devices. Layers of nitride 13 and HDP oxide 14 have been deposited over the gate regions and planarized, so that the gate regions 11, 12 are exposed. (In these embodiments of the invention, oxide region 14 is preferably HDP oxide rather than BPSG, in order to permit processing at lower temperatures.) Figure 1B is a longitudinal cross-

section view, showing that the dummy gate regions 11, 12 are formed on a sacrificial gate oxide layer 15 overlying substrate 1. Figure 1C is a transverse cross-section view, showing the nitride 13 and HDP oxide 14 on either side of the dummy polysilicon gate.

[0030] After the dummy gate and sacrificial gate oxide are removed, a layer of metal (thick enough to provide the desired work function) is deposited in the resulting trench, and the trench is filled with a suitable material such as undoped polysilicon. A polysilicon fill may then be converted to a silicide to provide a contact to the metal gate.

[0031] The metal gate structure may be either a single-metal gate with a midgap work function or a dual metal gate with different work functions for the n+ and p+ gate regions, as detailed below.

[0032] First Embodiment: Single-metal replacement gate

[0033] In this embodiment, the dummy polysilicon gate structures 11, 12 are removed in the same process; any convenient process having selectivity to nitride 13 and oxide 14 may be used. This process exposes the sacrificial oxide layer 15, which is then also removed. A trench 20 is thus formed with sidewalls 13a, 13b of nitride and an exposed portion of substrate 1 on the bottom. A new gate dielec-

tric 25 is formed on the exposed substrate, either by a deposition process or a thermal growth process. A conformal layer of metal 26 is then deposited as a blanket layer, covering the sidewalls of the trench and the gate dielectric 25. The composition and thickness of layer 26 are chosen to provide the desired work function, as is understood by those skilled in the art. Figures 2A and 2B are longitudinal and transverse cross-section views, respectively, of the structure after metal layer 26 is deposited.

[0034] A blanket layer 31 of undoped polycrystalline or amorphous silicon is then deposited over layer 26, with sufficient thickness to fill trench 20. The internal composition of layer 31 is chosen to provide the optimum trench fill; for example, it has been found that a mixture of polysilicon and amorphous silicon has desirable fill properties. Layers 31 and 26 are then planarized, preferably by CMP, so that the oxide region 14 is again exposed and the top surface 31a of the trench fill material is coplanar therewith. The resulting structure is shown in Figure 3A (longitudinal cross-section) and Figure 3B (transverse cross-section).

[0035] A blanket layer 41 of metal, suitable for forming a silicide, is then deposited (Figures 4A, 4B). This metal may be Co,

Ta, W, Mo or (preferably) Ni. A silicidation process is then performed (details of which are known in the art), so that the trench fill material is converted to a silicide 51. Another planarization process is performed to again expose the original top surface of the structure (see Figures 5A and 5B). The dummy gate structure is thus replaced by a single-metal gate with a silicide contact thereto (compare Figures 1C and 5B).

[0036] It will be appreciated that this process does not require masking of any layer, and is accordingly easy to implement.

[0037] Second Embodiment: Dual-metal replacement gate

[0038] In this embodiment, the dummy p+ and n+ polysilicon gate structures are removed separately, which requires masking of the polysilicon structure not being removed. Figure 6 shows the structure after masking the p+ dummy gate 11, removing the n+ dummy gate 12 and the underlying sacrificial gate oxide 15, and stripping the mask. A portion of the substrate 1 is thereby exposed. A blanket gate dielectric layer 75 is then grown or deposited, and a blanket metal layer 75 is deposited thereover (see Figure 7).

[0039] A blanket layer 81of undoped polycrystalline or amor-

phous silicon is then deposited over layer 26. As in the first embodiment, the composition of layer 81 is chosen to provide the optimum trench fill. The structure is then planarized to again expose dummy gate 11, so that the top surface 81a of the fill material 81 is coplanar therewith. The planarization process may be performed in a single step, e.g. CMP of the poly/amorphous silicon, continuing with CMP of the metal layer 76 endpointing on the surface of poly-Si gate 11. Alternatively, the planarization process may be performed in two steps, e.g. CMP of the poly/amorphous silicon endpointing on metal layer 76 overlying poly-Si gate 11, followed by a separate metal removal process. The resulting structure, after planarization, is shown in Figure 8.

[0040] The structure is then masked to permit removal of the dummy p+ polysilicon gate 11. The dummy p+ polysilicon gate 11 is removed along with the remaining sacrificial gate oxide 15. During the removal of the sacrificial oxide, a portion of the gate dielectric layer 75 covering the sidewall 76s of the metal is also removed. Figure 9 shows the structure after these steps and after the mask is removed; note that the metal sidewall 76s is exposed.

[0041] An oxidation process is then performed on the exposed

silicon surfaces, so that an oxide layer 85 is formed on the exposed portion of the substrate and an oxide 86 is formed on the surface of poly/amorphous silicon layer 81. The oxidation process typically includes a rapid thermal anneal (RTA) to about 900 °C, so that an oxide layer 95 is formed on the exposed surfaces of metal layer 76 (see Figure 10A). The gate dielectric layer 88 underlying the dual-metal gate is thus composed of layers 75 and 85, formed in separate steps.

[0042] A second blanket metal layer 96 is then deposited, as shown in Figure 10B. As in the previous steps, this blanket metal layer has a layer 91 of undoped polycrystalline or amorphous silicon deposited thereon. The structure is then planarized to yield the structure shown in Figure 11A (longitudinal cross-section) and Figure 11B (transverse cross-section). A comparison of Figures 11A and 11B with Figures 1B and 1C, respectively, shows that the dummy gate and sacrificial gate oxide have been replaced with a dual-metal gate with an underlying gate dielectric.

[0043] An etchback process is then performed which causes the gate region to be recessed with respect to the nitride region 13 and oxide region 14. Specifically, silicon layers 81, 91 are recessed relative to the oxide 14, and metal

oxide 95 and metal layers 76, 96 are recessed relative to the silicon layers 81, 91. A number of possible etch processes may be used, as is understood by those skilled in the art. The result of the etchback process is shown in Figure 12A (longitudinal cross-section) and Figure 12B (transverse cross-section).

[0044] A blanket layer of polycrystalline silicon is then deposited, so that a continuous silicon layer 98 overlies the dual-metal gate (Figure 13). This layer is planarized by RIE or CMP (or a combination thereof) so that nitride layer 13 and oxide layer 14 are again exposed. The resulting structure is shown in Figure 14A (longitudinal cross-section) and Figure 14B (transverse cross-section). At this point the dual-metal gate is recessed beneath a region of silicon whose top surface is coplanar with the surrounding dielectric materials (Figure 14B; compare with Figure 11B).

[0045] In this embodiment, a dual-metal gate structure is formed with a silicide contact thereto. It is advantageous to also provide a nitride cap layer for the structure, which permits annealing without affecting the metal or silicide. (In a preferred embodiment, the silicide is $NiSi_2$ and the annealing is done in forming gas. With a nitride cap, an anneal in forming gas at 400 °C may be used to remove excess

charge between the metal gate and the gate oxide, thereby preventing a shift in threshold voltage.) A nitride cap layer 120 may be deposited as a blanket layer, as shown in Figure 17A (longitudinal cross-section) and Figure 17B (transverse cross-section). Alternatively, the silicide layer 110 may be recessed (Figure 18A; compare Figure 16B), so that nitride layer 120 fills the recess (Figure 18B). Nitride layer 120 may then be planarized so that only the portion thereof overlying the silicide contact remains (Figure 18C). The nitride cap is then self-aligned to the dual-metal gate and silicide contact.

[0046] While the invention has been described in terms of specific embodiments, it is evident in view of the foregoing description that numerous alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, the invention is intended to encompass all such alternatives, modifications and variations which fall within the scope and spirit of the invention and the following claims.

[0047] We claim: